ORIGINAL ARTICLE 10.1111/j.1469-0691.2007.01808.x

Identification of *Legionella* spp. by 19 European reference laboratories: results of the European Working Group for Legionella Infections External Quality Assessment Scheme using DNA sequencing of the macrophage infectivity potentiator gene and dedicated online tools

*N. K. Fry[1], B. Afshar[1], W. Bellamy[1,2], A. P. Underwood[2], R. M. Ratcliff[3], T. G. Harrison[1] and members of the European Working Group for Legionella Infections**

[1]Respiratory and Systemic Infection Laboratory, [2]Statistics, Modelling and Bioinformatics Department, Health Protection Agency Centre for Infections, London, UK and [3]Infectious Diseases Laboratories, Institute of Medical and Veterinary Science, Adelaide, Australia

ABSTRACT

Identification of *Legionella* spp. can be achieved by DNA sequencing of the macrophage infectivity potentiator (*mip*) gene. The External Quality Assurance (EQA) scheme described in this report is the first to assess the proficiency of laboratories using this methodology. The results obtained from two EQA distributions sent to European reference laboratories involved in *Legionella* outbreak control and environmental monitoring are presented. Each distribution contained a panel of ten coded *Legionella* strains. All strains were from clinical and environmental sources and were considered to be wild-type strains. Participants used dedicated online tools to compare sequence text files against a database of known *Legionella* spp. The majority of centres (seven of ten, and 11 of 12) correctly identified all strains tested, in the first and second distributions, respectively. Typically, sequence similarity values of 98–100% were obtained when the test strains were compared with sequences contained in the database. In all but one case, lower values indicated a poor quality sequence. The exception was associated with the identification of a putative new species in the first panel. Genotypic identification of *Legionella* can be achieved by the use of standard protocols, dedicated identification libraries, and online tools. EQA schemes provide an independent measure of performance, and it is recommended that laboratories performing these techniques participate in such schemes, thereby allowing optimisation of and improvements in their performance.

Keywords DNA sequencing, external quality assessment, identification, *Legionella* spp., macrophage infectivity potentiator, online tools

Original Submission: 21 March 2007; **Revised Submission:** 24 May 2007; **Accepted:** 17 June 2007

Clin Microbiol Infect 2007; **13**: 1119–1124

Corresponding author and reprint requests: N. K. Fry, Respiratory and Systemic Infection Laboratory, Health Protection Agency Centre for Infections, 61 Colindale Avenue, London NW9 5EQ, UK
E-mail: Norman.Fry@HPA.org.uk
*Members of the European Working Group for Legionella Infections who participated in this study were as follows: J. Bangsborg, S. Blanco, J. Etienne, F. Fendukly, L. Franzin, V. Gaia, B. Herrmann, D. Lindsay, P. C. Lück, N. Maes, T. Marques, A. Papoutsi, C. Pelaz Antolín, M. Scaturro, A. van der Zee, P. Visca and G. Wewalka.

INTRODUCTION

To date, 53 species of *Legionella* have been described [1–3], including one genomospecies [4] and one *Candidatus* species [2]. Approximately one-third of these have been associated with human disease following isolation of the organism from clinically significant infections [1]. The identification of the majority of these species using classical phenotypic methods is very difficult. Several

genotypic methods have been described, including analysis of total genomic DNA (e.g., ribotyping and random amplified polymorphic DNA analysis [5,6]), analysis of specific genes (e.g., rRNA [7], macrophage infectivity potentiator (*mip*) [8], RNA polymerase (*rpoB*) [9], RNase P RNA (*rnpB*) [10] and *gyrA* [11] genes), and analysis of internal regions of these genes (e.g., intergenic regions of rDNA [12] or regions coding for tRNA [13]).

Of these various approaches, PCR amplification and sequencing of the *mip* gene [8] appears to be the best method, for three main reasons. First, the *mip* sequence has more informative sites, both in length and percentage of available bases, than some of the other genes listed above, e.g., 16S rDNA [14]. Second, the procedure is easy to perform, involving only a simple primary PCR and a single sequencing reaction. Third, a comprehensive, dedicated and freely accessible web database has been established by members of the European Working Group for Legionella Infections (EWGLI). This contains sequence data from all validly described species, and many others awaiting formal description, and allows the online identification of putative *Legionella* spp.

The aim of the External Quality Assessment (EQA) scheme, devised by EWGLI members at the Health Protection Agency Centre for Infections (HPA CfI), London, UK, was to investigate the ability of European microbiology reference laboratories to identify coded panels of *Legionella*, using a standard protocol and dedicated online tools. This report presents the results from the first two EQA distributions.

MATERIALS AND METHODS

Participants

Participation was open to any EWGLI collaborating laboratory. Overall, 19 institutions from 14 countries took part in one or both distributions. Each of these centres acts as a local or national reference laboratory for *Legionella* infections. The study was coordinated on behalf of EWGLI by the HPA CfI. The participating laboratories were: Austrian Agency for Health and Food Safety, Institute of Medical Microbiology and Hygiene, Vienna, Austria; Laboratoire de Référence *Legionella*, Université Libre de Bruxelles-Hôpital Erasme, Brussels, Belgium; Herlev Hospital, Herlev, Denmark; Centre National de Référence des Légionelles, Lyon, France; Technical University Dresden, Dresden, Germany; Aristotle University of Thessaloniki, Thessaloniki, Greece; Institute for Infectious Diseases, IRCCS Lazzaro Spallanzani, Rome, Italy; Istituto Superiore di Sanità, Rome, Italy; Amedeo di Savoia Hospital, Turin, Italy; St Elisabeth Hospital, Tilburg, The Netherlands; Hospital de Santa Cruz, Carnaxide, Portugal; Scottish *Legionella* Reference Laboratory, Glasgow, Scotland; Hospital Universitari Germans Trias i Pujol, Badalona, Spain; Instituto de Salud Carlos III, Madrid, Spain; Karolinska University Hospital, Stockholm, Sweden; University Hospital, Uppsala, Sweden; Istituto Cantonale di Microbiologia, Bellinzona, Switzerland; and HPA CfI.

Study design

The study was designed in two phases. First, a panel of ten strains (see below) was distributed to participants (Distribution 1). Following analysis of the results submitted, and reports back to participants, the procedures were reviewed, and training needs were identified. Following the provision of necessary training, a second panel was distributed (Distribution 2) and the participants' ability was reassessed. Oligonucleotides for PCR amplification and DNA sequencing were not provided by the coordinating centre, but participants were asked to provide information concerning the sequencing platform used.

Preparation of proficiency panels

Both proficiency panels contained ten coded *Legionella* strains: Panel 1, strains A–J, and Panel 2, strains K–T (Table 1); three strains were included in both panels. With one exception, the strains had been fully characterised previously by phenotypic, serological and genotypic methods. The exception was the strain coded G, which had only been characterised before its inclusion in the panel as being serologically closest to *Legionella moravica*. According to *mip* gene sequencing, this strain also showed the highest similarity to *L. moravica*, but at a much lower level than the other strains (Table 1). All isolates are held in the Respiratory and Systemic Infection Laboratory (RSIL; HPA CfI) or EWGLI culture collections, and are considered to be wild-type strains (i.e., low-passage non-reference strains).

Table 1. Strains of *Legionella*, representing 11 species, used in the proficiency panels

Legionella species	Strain identifier[a]	Source[b]	Study code identifier	Sequence similarity (%) to *mip* profile on database[c]
L. anisa	LC3967	E	A	100
L. bozemanii	LC2915	C	B	100
L. dumoffi	LC5207	E	C/N	100
L. jamestowniensis	LC5389	E	D/R	100
L. jordanis	LC5073	E	E	99.1
L. longbeachae	LC4987	C	F	100
Legionella sp.[d]	LC5212	E	G	94.6
L. pneumophila	LC6635	C	H	100
L. quinlivanii	LC5390	E	I	100
L. sainthelensi	LC5068	E	J/T	100
L. bozemanii	LC4348	C	K	100
L. longbeachae	LC4790	C	L	100
L. jordanis	LC3940	E	M	100
L. bozemanii	LC4348	C	O	100
L. pneumophila	H034700617	C	P	99.6
L. oakridgensis	LC 3780	E	Q	100
L. pneumophila	EUL 137	E	S	100

[a]Strains are held in the Respiratory and Systematic Infection Laboratory, Health Protection Agency Centre for Infections, London, UK (LC and H0 prefix) or European Working Group for Legionella Infections culture collection (EUL prefix).
[b]C, clinical isolate, E, environmental isolate.
[c]Results from forward and reverse sequencing reactions (or forward on more than one occasion) were determined by the coordinating laboratory before despatch.
[d]This strain was initially characterised as closest serologically, and by *mip* gene sequencing, to *L. moravica*, but was subsequently shown to be a new species of *Legionella*.

Replicates of each strain were prepared on buffered charcoal yeast extract (BCYE) agar (Oxoid, Basingstoke, UK) slopes, and were then distributed to participants by courier. Panel 1 was dispatched on 13 August 2003 and Panel 2 was dispatched on 8 October 2004.

Specimen processing

Upon receipt, each centre subcultured the coded strains on BCYE plates according to standard methods [15]. Failure to recover a strain was scored as 'no growth', and these results were excluded from the final analysis. Each centre performed genomic DNA extraction according to their current methodology. The method of Ratcliff *et al*. [8] was used for species identification. Primary amplification was performed using the primers Legmip_f 5′-GGG(A/G)ATT(A/C/G)TTTATGAAGATGA(A/G)A(C/T)TGG and Legmip_r 5′-TC(A/G)TT(A/T/C/G)GG(A/T/G)CC(A/T/G)AT(A/T/C/G)GG(A/T/C/G)CC(ATG)CC. DNA sequencing was performed using the primer Legmip_fs 5′-TTTATGAAGATGA(A/G)A(C/T)TGGTC(A/G)CTGC according to local practice. The resulting DNA sequence was entered into the database (http://www.hpa-bioinfotools.org.uk/mip_ID.html) as a flat text file in order to allow identification using the similarity tool. The species revealed by the similarity tool as showing the highest percentage similarity to the sequence obtained was taken to be the species identification. The species and similarity (%) were reported for each strain tested.

Scoring

The performance of each centre was measured in two ways. The first measure was the ability to correctly identify the coded isolates against the coordinating laboratory's intended identification (intended result). For the purpose of these distributions, identifications based on sequence similarity values of ≤80% were considered to be unacceptable and were scored as 'not identified'. Second, laboratories were scored according to the confidence in the identification made, i.e., identification was considered to be 'confident' or 'tentative'. For a confident identification, the *mip* sequence submitted had to show ≥98.0% similarity to a sequence in the database, while submissions of 80–98% similarity were scored as tentative. A combined score of 'tentative' plus 'confident', where the correct species identification was obtained, was also calculated. Results for Panel 1 were only scored for nine strains, as strain G was excluded from the analysis.

Software

Access to the *Legionella mip* gene sequence database is via the EWGLI website (http://www.ewgli.org/) or the HPA bioinformatics webpages (http://www.hpa.org.uk/cfi/bioinformatics/ewgli/legionellamips.htm). Details of 157 *Legionella* strains provided by one of the authors (RMR) [16] are stored in a PostgreSQL database. The associated *mip* sequences are stored both in this database and in a local Basic Local Alignment Search Tool (BLAST) [17] database. Software was written to provide a web-based graphical interface to this database, predominantly using perl-cgi and BioPerl Modules [18]. Upon submission of a text sequence to the form on the web page, the sequence is used as a query in a blastn search against the *mip* sequence BLAST database. The full sequences of the top five database matches are retrieved and aligned with the sequence submitted by the user and a set of *mip* reference sequences selected to cover the entire range of the *Legionella* genus. The alignment is performed using the MAFFT algorithm [19]. Using this procedure, the online tool and underlying database provide the following functions: (i) an alignment of all the sequences from the reference alignment, top five database matches and the user sequence; (ii) a neighbour-joining tree of the alignment, including the reference species, five closest matches from the database and the user sequence; and (iii) an alignment of the eight sequences from the combined alignment that are most similar to the user sequence, together with their percentage similarity scores.

RESULTS

Distribution 1

The first panel was distributed to 16 centres in nine countries. Results were returned by ten centres before the deadline (Table 2). Four centres did not recover isolate J successfully, and not all centres reported results for all isolates. Strain G was found by all ten laboratories to have <98% similarity to any sequence in the database, although it showed most similarity (*c*. 95%) to *L. moravica*. Seven of the ten centres identified correctly all strains tested. Two centres misidentified one strain, and one centre misidentified (or failed to identify) five of the seven strains examined. All of the identifications were 'confident' identifications in seven of the laboratories; results from the other centres were mainly 'tentative' or <80%, suggesting that the quality of the *mip* sequences was poor. Following analysis and feedback to the participants, training needs were highlighted and the coordinating centre provided training to three centres before distribution of the second panel.

In the case of the *Legionella jordanis* strain, the coordinating centre determined 560 nucleotides of the *mip* gene before dispatch of the distribution, which showed a maximum similarity value of 99.1% (555/560 nucleotides) to the reference sequence in the database, i.e., five nucleotide differences. Most of these differences were located at the 3′-end of the sequence; thus, the most likely explanation for participants scoring above this value was the submission of shorter sequences, i.e., <560 nucleotides.

Distribution 2

The second panel was distributed to 19 centres in 11 countries. Valid results were returned by 12

Table 2. Results obtained with the first *Legionella* distribution[a]

Study code no.	Intended result	Species	Centre no. 2	4	6	9	11	13	16	19	20	15	No. achieving correct species identification
A	*L. anisa*		✔	94.0	✔	*L. dumoffii* (99.8)	✔	✔	✔	✔	92.6	Not identified (<80)	8/10
B	*L. bozemanii*		✔	92.1	✔	✔	NT	✔	✔	✔	97.8	Not identified (<80)	8/9
C	*L. dumoffii*		✔	*L. anisa* (94.0)	✔	99.6	NT	✔	✔	✔	98.0	*L. bozemanii* (99.8)	7/9
D	*L. jamestowniensis*		99.8	83.0	✔	99.7	NT	99.8	✔	✔	NT	*L. dumoffii* (100)	7/8
E	*L. jordanis*		✔	94.6	98.0	✔	99.5	✔	✔	✔	97.7	98.8	10/10
F	*L. longbeachae*		✔	98.4	✔	98.7	✔	✔	✔	✔	98.3	*L. jordanis* (96.7)	9/10
G[b]	*'L. moravica'*		94.6	82.9	94.6	94.6	96.0	94.4	94.5	94.6	95.7	96.0	NA
H	*L. pneumophila*		99.6	94.9	98.8	✔	99.8	✔	✔	✔	97.4	*L. birminghamensis* (92.3)	9/10
I	*L. quinlivani*		✔	88.9	✔	✔	NT	99.5	99.5	✔	NT	*L. pneumophila* (97.6)	7/8
J	*L. sainthelensi*		✔	NG	98.2	✔	NT	✔	NG	NG	97.1	NG	5/5
Centre score	Intended identification	Tentative	0/9	6/8	0/9	0/9	0/4	0/9	0/8	0/8	5/7	0/8	
		Confident	9/9	1/8	9/9	8/9	4/4	9/9	8/8	8/8	2/7	1/8	
		Combined score	9/9	7/8	9/9	8/9	4/4	9/9	8/8	8/8	7/7	1/8	

NT, not tested; NA, not applicable; NG, no growth.
[a]Sequence similarity values of 100% are indicated by '✔'. Values ≥98.0% concordant with the intended identification were scored as 'confident'; values of 80–98% were scored as tentative; and values ≤80% were scored as 'not identified'.
[b]Results for isolate G were not included in the analysis.

centres before the deadline, including seven that had submitted results for Distribution 1 (Table 3). All but one centre correctly identified all strains tested, and this centre only failed to identify one strain. Furthermore, only two centres reported identifications that were not considered 'confident'. One centre, which had sent personnel for training at HPA CfI, achieved an improved score.

DNA sequencing platforms

In the first distribution, seven of ten centres used ABI platforms, i.e., ABI Prism 310, ABI 377, ABI 3100, ABI 3700 or ABI 3730 Genetic Analyzers (Applied Biosystems, Foster City, CA, USA). The remainder used either the CEQ8000 Genetic Analysis System (Beckman Coulter, Fullerton, CA, USA) or the MegaBACE DNA Sequencing Platform (Amersham Biosciences, Piscataway, NJ, USA). Similarly, in the second distribution, ten of the 12 participants used one of the ABI platforms listed above, one used the CEQ8000 Genetic Analysis System, and one used the MegaBACE DNA Sequencing Platform. There was no apparent difference in the sequence quality obtained with the different platforms.

Table 3. Results obtained with the second *Legionella* distribution[a]

Study code	Intended result		Centre no. 2	6	8	9	11	13	15	17	18	19	20	21	No. achieving correct species identification
K	*L. bozemanii*		✔	✔	✔	✔	96.9	✔	80.9	✔	✔	99.8	✔	✔	12/12
L	*L. longbeachae*		✔	✔	✔	✔	85.1	✔	96.4	✔	✔	✔	✔	✔	12/12
M	*L. jordanis*		✔	99.8	✔	✔	95.2	✔	97.2	NT	✔	99.8	✔	✔	11/11
N	*L. dumoffii*		✔	✔	99.5	✔	97.5	✔	98.9	✔	✔	99.3	99.8	✔	12/12
O	*L. bozemanii*		✔	✔	✔	✔	95.5	99.8	99.1	✔	✔	98.9	✔	✔	12/12
P	*L. pneumophila*		✔	99.5	✔	99.7	NT	99.5	90.5	98.7	✔	98.6	98.0	✔	11/11
Q	*L. oakridgensis*		✔	✔	99.5	✔	94.4	✔	99.6	✔	✔	✔	99.6	✔	12/12
R	*L. jamestowniensis*		99.5	99.8	99.4	99.6	Not identified (<80)	✔	99.3	NT	NT	✔	98.5	✔	9/10
S	*L. pneumophila*		✔	✔	99.6	✔	NT	✔	✔	99.1	✔	✔	✔	✔	11/11
T	*L. sainthelensis*		✔	✔	✔	✔	95.0	✔	98.9	✔	✔	NT	99.8	✔	11/11
Centre score	Intended identification	Tentative	0/10	0/10	0/10	0/10	7/8	0/10	4/10	0/8	0/9	0/9	0/10	0/10	
		Confident	10/10	10/10	10/10	10/10	0/8	10/10	6/10	8/8	9/9	9/9	10/10	10/10	
		Combined score	10/10	10/10	10/10	10/10	7/8	10/10	10/10	8/8	9/9	9/9	10/10	10/10	

NT, not tested.
[a]Sequence similarity values of 100% obtained using *mip* are indicated by '✔'. Values ≥98.0% concordant with the intended identification were scored as 'confident'; values of 80–98% were scored as 'tentative'; and values ≤80% were scored as 'not identified'.

DISCUSSION

It is now clear that genotypic methods are essential in order to identify all species of *Legionella*. Several such methods have been described, but sequencing of the *mip* gene currently appears to provide the best option. With increased use of DNA sequencing in microbiological laboratories, EQA schemes to monitor such techniques are essential to ensure competence. To our knowledge, the EQA scheme described in this report is the first such large-scale study to evaluate genotypic methods for bacterial species identification. The results obtained show clearly that identification of *Legionella* spp. by *mip* gene sequencing works well, and that most reference laboratories can get the correct (intended) answer. While it is acknowledged that best practice would be to include only well-characterised isolates in EQA panels, the mistake made in the inclusion of strain G served to illustrate that putative novel species can also be recognised using this method. The study also revealed that training in DNA sequencing is sometimes necessary, and that good quality training and increased experience lead to higher quality and more reliable results.

No single system is perfect, and it is acknowledged that one species, *Legionella geestiana*, is known not to amplify with the primers described above. Additional primers capable of amplifying the *mip* gene from *L. geestiana* have been described (e.g., 5′-GTNACNGTNGANTANAC-NGG together with Legmip_r) [16], or another gene target, e.g., *rpoB* or *rnpB*, can be used [9,10]. Improvements to the current scheme are planned, in both the practical methodology and the analytical strategy. The reverse amplification primer described in the original method does not generate acceptable quality sequences consistently from all species; however, alternative strategies can be devised to circumvent this problem (B. Wullings, personal communication).

Currently, identification of an unknown strain with a species present in the database should provide a similarity score of 98–100% with a good-quality sequence. It is notable that although the majority of wild-type strains in the panels (17/20) had *mip* sequences identical to those of a strain contained in the database, the maximum similarity possible (100%) was clearly not obtained by the majority of participants; however, most correct identifications showed >98% similarity. This suggests that the sequence text file used for identification contained errors or ambiguities, most likely following submission of an unedited or poorly edited sequence. In response to these studies, an automated sequence quality tool has been implemented, and this now helps to prevent such misidentification. Single chromatogram files can be uploaded directly, and the automated tool provides feedback on the quality of the submitted trace file, based on the phred algorithm [20,21], thus making the result less subject to interpretive bias, although a minimum length of 300 nucleotides is still recommended. The ability to upload forward and reverse sequence trace files is planned. As with all similar databases, there is also a requirement for active curation in order to add new profiles and information.

Genotypic identification of *Legionella* spp. is an essential requirement for reference laboratories. Standard protocols, dedicated identification libraries and online tools are valuable resources to help achieve this goal. It is anticipated that additional genes, including those coding for 16S rRNA, RpoB, RnpB and GroEL, will be added to the current identification system to aid in the polyphasic characterisation and identification of known and potential novel members of this genus. In addition, it is intended to augment the *Legionella* spp. *mip* database to include representatives of species with different sequence types (e.g., as above), together with relevant epidemiological information. Contribution of novel sequences or information relating to existing sequences in different geographical regions is welcomed by the curators.

ACKNOWLEDGEMENTS

We thank R. George and M. Struelens for critical reading of this manuscript. This work was supported, in part, by funding from the European Centre for Disease Prevention and Control (contract number ECD 368). Preliminary data from this study were presented at the 6th International Conference on *Legionella* (Chicago, IL, USA).

REFERENCES

1. Harrison TG. *Legionella*. In: Borriello SP, Murray PR, Funke G, eds, *Topley & Wilson's microbiology & microbial infections*. London: Hodder Arnold, 2005; 1761–1785.
2. Park M, Yun ST, Kim MS, Chun J, Ahn TI. Phylogenetic characterization of *Legionella*-like endosymbiotic X-bacteria

in *Amoeba proteus*: a proposal for 'Candidatus *Legionella jeonii*' sp. nov. *Environ Microbiol* 2004; **6**: 1252–1263.
3. Kuroki H, Miyamoto H, Fukuda K *et al. Legionella impletisoli* sp. nov. and *Legionella yabuuchiae* sp. nov., isolated from soils contaminated with industrial wastes in Japan. *Syst Appl Microbiol* 2007; **30**: 273–279.
4. Benson RF, Thacker WL, Daneshvar MI, Brenner DJ. *Legionella waltersii* sp. nov. and an unnamed *Legionella* genomospecies isolated from water in Australia. *Int J Syst Bacteriol* 1996; **46**: 631–634.
5. Salloum G, Meugnier H, Reyrolle M *et al*. Identification of *Legionella* species by ribotyping and other molecular methods. *Res Microbiol* 2002; **153**: 679–686.
6. Van Belkum A, Struelens M, Quint W. Typing of *Legionella pneumophila* strains by polymerase chain reaction-mediated DNA fingerprinting. *J Clin Microbiol* 1993; **31**: 2198–2200.
7. Fry NK, Warwick S, Saunders NA, Embley TM. The use of 16S ribosomal RNA analyses to investigate the phylogeny of the family *Legionellaceae*. *J Gen Microbiol* 1991; **137**: 1215–1222.
8. Ratcliff RM, Lanser JA, Manning PA, Heuzenroeder MW. Sequence-based classification scheme for the genus *Legionella* targeting the *mip* gene. *J Clin Microbiol* 1998; **36**: 1560–1567.
9. Ko KS, Lee HK, Park M-Y *et al.* Application of RNA polymerase β-subunit gene (*rpoB*) sequences for the molecular differentiation of *Legionella* species. *J Clin Microbiol* 2002; **40**: 2653–2658.
10. Rubin C-J, Thollesson M, Kirsebom LA, Herrmann B. Phylogenetic relationships and species differentiation of 39 *Legionella* species by sequence determination of the RNase P RNA gene *rnpB*. *Int J Syst Evol Microbiol* 2005; **55**: 2039–2049.
11. Feddersen A, Meyer HG, Matthes P, Bhakdi S, Husmann M. GyrA sequence-based typing of *Legionella*. *Med Microbiol Immunol* 2000; **189**: 7–11.
12. Liu H, Li Y, Huang X, Kawamura Y, Ezaki T. Use of the *dnaJ* gene for the detection and identification of all *Legionella pneumophila* serogroups and description of the primers used to detect 16S rDNA gene sequences of major members of the genus *Legionella*. *Microbiol Immunol* 2003; **47**: 859–869.
13. De Gheldre Y, Maes N, Presti FL, Etienne J, Struelens M. Rapid identification of clinically relevant *Legionella* spp. by analysis of transfer DNA intergenic spacer length polymorphism. *J Clin Microbiol* 2001; **39**: 162–169.
14. Ratcliff RM. The problem of complexity. In: Cianciotto NP, Abu Kwaik Y, Edelstein PH *et al*., eds, *Legionella: state of the art 30 years after its recognition*. Washington, DC: ASM Press, 2006; 359–366.
15. Harrison TG, Taylor AG, eds. *A laboratory manual for Legionella*. Chichester: Wiley, 1988.
16. Ratcliff RM. Genetic and functional studies of the Mip protein of *Legionella*. PhD Thesis, University of Adelaide, 2000.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.
18. Stajich JE, Block D, Boulez K *et al*. The Bioperl Toolkit: perl modules for the life sciences. *Genome Res* 2002; **12**: 1611–1618.
19. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002; **30**: 3059–3066.
20. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using *phred*. I. Accuracy assessment. *Genome Res* 1998; **8**: 175–185.
21. Ewing B, Green P. Base-calling of automated sequencer traces using *phred*. II. Error probabilities. *Genome Res* 1998; **8**: 186–194.